

Transcrição

Neste passo, falaremos sobre a medida **moda**.

Esta é a mais fácil de definir entre as demais Medidas de Tendência Central, pois é justamente o **valor mais frequente** dentro de um conjunto de dados.

Na primeira célula da parte "3.3 Moda" de nosso notebook, teremos o mesmo `df` com as notas de três alunos em sete matérias.

`df`

Matérias	Beltrano	Fulano	Sicrano
Matemática	10.0	8	7.5
Português	2.0	10	8.0
Inglês	0.5	4	7.0
Geografia	1.0	8	8.0
História	3.0	6	8.0
Física	9.5	10	8.5
Química	10.0	8	7.0

Observando somente as notas de Beltrano, é evidente que a moda com o valor que mais aparece é a "10.0", pois temos um pequeno conjunto de dados.

É possível termos mais de uma moda dentro em um dataset; por exemplo, se houver apenas duas, chamaremos a variável como **bimodal**. Caso tenha

Na célula seguinte, obteremos a medida com o método `.mode()` da biblioteca Pandas.

`df.mode()`

Matérias	Beltrano	Fulano	Sicrano
0	10.0	8	8.0

No retorno, veremos um `Dataframe` com os **valores modais**.

Para termos um exemplo de variáveis multimodais, criaremos a `exemplo` sendo igual a uma nova `pd.Series()` contendo uma lista de valores.

`exemplo = pd.Series([1, 2, 2, 3, 4, 4, 5, 6, 6])`

Com isso, veremos a mesma quantidade de ocorrências dos valores 2, 4 e 6. Chamando somente `exemplo` na sequência, veremos o seguinte res

```
0    1
1    2
2    2
3    3
4    4
```

```
5    4
6    5
7    6
8    6
```

```
dtype: int64
```

Obtendo este índice com os valores, aplicaremos `exemplo` com `.mode()`.

```
exemplo.mode()
```

```
0    2
1    4
2    6
```

```
dtype: int64
```

O resultado será uma *series* com os três valores modais `2`, `4` e `6`, provando que a variável é multimodal de fato. Se alterássemos o último valor p

Portanto, esta medida apresenta os valores que são mais frequentes em um conjunto de dados.

Este tipo de Estatística Descritiva é bastante utilizada com variáveis qualitativas que possuem valores categóricos, pois as contínuas podem ter inúmer

Em "Obtendo a moda em nosso dataset", escreveremos `dados` com a variável `Renda` seguida de `mode()` apenas para avaliarmos o retorno.

```
dados.Renda.mode()
```

Como resultado, veremos o valor `788` que corresponde ao salário mínimo no ano de 2015 quando a PNAD foi realizada. Como já vimos, a grande r
sentido que este número ocorra com mais frequência.

Na célula seguinte, pegaremos a moda da variável `Altura` que resultará na seguinte saída:

```
0    1.568128
1    1.671225
2    1.681659
3    1.692977
4    1.708163
5    1.708370
6    1.753842
7    1.779073
8    1.796462
```

```
dtype: float64
```

Logo, veremos que este elemento é multimodal e possui nove modas no conjunto de valores. Também veremos que esta medida para um valor contém

A seguir, entenderemos as relações entre as três estatísticas descritivas média, mediana e moda quando conseguirmos identificar **simetrias** e potências